The Integrated Exposure Uptake Biokinetic Model for Lead in Children: Independent Validation and Verification

Larry Zaragoza[1] and Karen Hogan[2]

[1]Office of Solid Waste and Emergency Response, [2]Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, Washington, DC

The U.S. Environmental Protection Agency employs a model, the integrated exposure biokinetic (IEUBK) model for lead in children, for the assessment of risks to children posed by environmental lead at hazardous waste sites. This paper describes results of an effort to verify the consistency of the documentation with the computer model and to test the computer code using a group that is independent from those involved in the model development. This review concluded that the IEUBK model correctly calculates the equations specified in the IEUBK model theory documentation. However, several issues were identified on model documentation, model performance, and the C++ programming language code (i.e., IEUBK model source code) documentation. These issues affect the ability of an independent reviewer to understand the workings of the IEUBK model but not the model's reliability. As a result of these findings, recommendations have been provided for updating documentation to the model as well as associated adjustments to the model documentation. — *Environ Health Perspect* 106(Suppl 6):1551–1556 (1998). *http://ehpnet1.niehs.nih.gov/docs/1998/Suppl-6/1551-1556zaragoza/abstract.html*

Key words: models, model validation, lead

As outlined in other papers in this monograph, the integrated exposure uptake biokinetic (IEUBK) model for lead in children (*1*) is a mathematical model that has been developed for integrating lead exposure across multiple exposure pathways for children. This model provides a multipathway analysis of the impact of environmental lead levels that relies upon site-specific information. Before adoption of the IEUBK as a tool to assess lead risks at sites, the U.S. Environmental Protection Agency (U.S. EPA) employed a range of 500 to 1000 ppm to support site cleanup decisions. In many instances, interpretation of this range resulted in the use of 500 ppm for residential cleanups and 1000 ppm for industrial cleanups. Today, the U.S. EPA's use of the IEUBK model encourages use of site-specific information for residential settings and often results in cleanup levels that are higher than the 500 ppm that would have been used to protect the health of children.

The IEUBK model relates environmental concentrations of lead with potential blood lead levels in children exposed to contaminated medium. The IEUBK model is structured so that the environmental concentration–blood lead relationship in children is established through four distinct components: exposure, uptake, biokinetics, and blood lead distributions. These four model components are designed to run as distinct but interrelated modules, and can be described as follows:

- The exposure components compute an exposure/intake–dose, expressed as micrograms of lead per day, based on media-specific lead concentrations and consumption rates (in cubic meters of air inhaled per day, grams of soil ingested per day, or liters of water ingested per day).
- The uptake component estimates the biologic uptake and transfer of lead from the gastrointestinal tract or lungs to the blood (in micrograms per day) for children ages 0 to 7 years (0–84 months).
- The biokinetic component estimates the transfer of absorbed lead between blood and other vital tissues and its elimination through excretory pathways. The outcomes are calculated in discrete time fractions for the period of 0 to 84 months.
- The probability distribution component produces graphic illustration of the probability of exceeding blood lead levels above the level of concern (default value of 10 µg/dl) for a particular age group (or time period) for up to 84 months. The user can then explore an array of possible changes in exposure media that would reduce the probability that blood lead concentrations would be above this level of concern.

The IEUBK model is a simulation model that should be viewed as a tool for making rapid calculations and recalculations of a complex set of equations that includes a large number of exposure, uptake, and biokinetic parameters. In addition to assessing childhood lead exposure and retention, the IEUBK model results can be a useful component of remediation strategies for lead in the environment.

The IEUBK model is a product of many years of development within the U.S. EPA. The initial efforts to model lead emerged from the Office of Air Quality Planning and Standards with the development of the National Ambient Air Quality Standard for Lead and subsequently from the Office of Water in the National Primary Drinking Water Regulation for Lead. Both of these offices employed mathematical modeling to estimate the impact of lead on child blood lead levels.

Lead is one of the most prevalent toxic chemicals found at Superfund sites. The U.S. EPA data show that lead is among the most frequently used contaminant in the scoring of sites with the Hazard Ranking System, which is the primary tool the U.S. EPA employs to add hazardous waste sites to the National Priorities List (NPL) (2). NPL sites are those hazardous waste sites that are determined to warrant raising their status to have national concerns. At present there are about 1200 proposed and final NPL sites across the United States.

One of the uses of this IEUBK model is to support the implementation of the 1994 interim directive of the Office of Solid Waste and Emergency Response (3) for the assessment of soil lead risks. The interim directive explains how the IEUBK model results can be used as a tool to assist in determining site-specific cleanup levels. In this context, the IEUBK model can be viewed as a predictive tool for estimating changes in blood lead concentrations as exposures are changed. Also, the IEUBK model could be viewed as a useful tool to aid the agency in making more informed choices about the concentrations of lead that might be expected to impact human health. However, it is important to recognize that the outputs of the IEUBK model alone do not determine cleanup levels; other factors as outlined in the National Contingency Plan are also considered in determining cleanup levels.

The need for the review described in this paper is detailed in the validation strategy for the IEUBK model (4). This validation strategy specifically calls for a four-step validation process. These steps are described below.

a) The scientific foundations of the model structure. Does the model adequately represent the biological and physical mechanisms of the modeled system? Are the mechanisms understood sufficiently to support modeling?

b) Adequacy of parameter estimates. How extensive and robust are the data used to estimate model parameters? Does the parameter estimation process require additional assumptions and approximations?

c) Verification. Are the mathematical relationships posited by the model correctly translated into computer code? Are model inputs free from numerical errors?

d) Empirical comparisons. What are the opportunities for comparison between

model predictions and data, particularly under conditions under which the model will be applied in assessments? Are model predictions in reasonable agreement with relevant experimental and observational data?

The work described below has been performed to address step c of the validation process.

Methodology

The IEUBK Independent Validation and Verification (IV&V) project was designed to verify that the computational algorithms can accurately solve the governing equations and parameters and that the code is fully operational (i.e., the code is robust and serviceable).

The IV&V team investigated other validation methodologies to determine whether the approach taken here is consistent with current practices. Although we were unable to locate validation studies for other models that specifically use differential equations, the general validation and verification process applied to the IEUBK model is consistent with software testing standards used. In particular, the standard software test process outlined advocates considering automated scripts or drivers, and interface simulators as tools for software testing. The Mathcad 5.0 software (5), which was used to reproduce the IEUBK model functions, is such a tool. In addition, standardized testing routines and reporting formats, as well as configuration control and problem logs, were used to track issues identified in the validation.

To conduct the IV&V of the IEUBK model, the following activities were planned and implemented:
• Review of IEUBK model documentation and other related materials for completeness and accuracy.
• Comparisons between the IEUBK model source code equations and parameters with the IEUBK model's theory.
• Verification of the IEUBK model source code, including coding of equations, for completeness and correctness.
• Examination of the efficiency of the IEUBK model source code.
• Review of the IEUBK model for compliance with agency standards for the development of scientific information systems.
• Pathway tests that evaluated the performance of the system on various hardware configurations.

The following discussion focuses on the results for the first three items described

above. The IV&V work undertaken was based on review of a copy of the source code, publicly available documentation, and a copy of the IEUBK model source code, which is commercially available. The U.S. EPA used a group to perform the work described below that was independent of the parties involved in the original development and programming of the IEUBK model.

Findings

The findings of these activities, summarized below, are drawn from the IV&V report (6). The details of the validation strategy and technical approach can be found in the IV&V report. The IV&V report concludes that the IEUBK model correctly calculates the equations specified in the IEUBK model theory. The IV&V report does, however, identify minor issues that do not affect the results of the model. These issues, presented in the following sections, relate to model documentation errors, model performance, and source code documentation.

Comparison of the List of Equations and Parameters with the IEUBK Model's Theory

The IV&V report documents the comparison of the equations provided in Appendix A of the *Technical Support Document: Parameters and Equations Used in the Integrated Exposure Uptake Biokinetic Model for Lead in Children, v 0.99d* (the TSD) (7) to the theory presented in the main body of that report and in the *Guidance Manual for the Integrated Exposure Uptake Biokinetic Model for Lead in Children (1)*. In addition, the IV&V report documents comparisons of the TSD to the *Correspondence Between the IEUBK Model Source Code and Technical Support Document: Parameters and Equations Used in the Integrated Exposure Uptake Biokinetic Model for Lead in Children, v 0.99d* (the Bridge document) (8). Table 1 provides a summary of the discrepancies found during the comparison.

Verification of the Default Values for the Parameters in the IEUBK Model Source Code

The report describes the verification of the coding of default parameter values in the IEUBK model source code. A summary of the results is presented in Table 2. None of the issues identified during this activity affect the results calculated by the IEUBK model.

Verification of the Execution of the Linear Equations in the Code

The execution of the linear equations for calculation of environmental pathway exposure and uptake of lead, and calculation of compartmental lead transfer times, fluid volumes, and organ weights are presented in Table 3.

Verification of the Execution of the Differential and Associated Equations in Isolation

The IV&V report documents the execution of the differential equations in the IEUBK model and their associated equations in the calculation of compartmental lead masses, tissue lead masses, and blood lead concentration at birth, and the overall blood lead concentrations in the TSD. A summary of the results of this activity is presented in Table 4.

Testing of the Overall Set of Equations

The overall set of equations in the IEUBK model source code was verified and the results are summarized in Table 5.

Verification of the Input/Output Routines

The IV&V report verified the input/output routines of the IEUBK model and documented all discrepancies in the input/output routines that control user-entered parameter values. A summary of the results is presented in Table 6.

Investigation of Geometric Mean Calculations

The IEUBK model source code for the calculation of the geometric mean was reviewed. Two functions were identified in the IEUBK model source code calculation of the geometric mean. These functions use a time-weighted average of monthly blood lead levels to provide a central tendency blood lead level corresponding to each year of age. This time-weighted average is algebraically equivalent to an arithmetic mean of the monthly predicted blood lead levels. The IV&V report noted an apparent conflict between an average calculated as an arithmetic mean, yet subsequently used as a geometric mean. The U.S. EPA considers the algorithm correct for the reasons discussed below. Any deficiency in the approach is due more to a lack of adequate documentation. Documentation will be clarified.

The IEUBK model generates a central tendency estimate of blood lead for a typical individual smoothed over each year of age up to 7 years. The time-weighted average blood lead level is the last step in the deterministic, biokinetic component of the model before the model estimates individual risk of elevated blood lead in its statistical component. The time-weighted average deals with variability over time in lead biokinetics. Given the current level of understanding of these processes, equal weighting of monthly central tendency estimates is the most defensible approach to estimating an overall central tendency blood lead level. There is no evidence that variability over time should be characterized by lognormal variability.

The statistical component of the IEUBK model then generates a lognormal distribution of individual blood lead levels around the best central tendency estimate of blood lead for a given lead exposure scenario. There is strong support for the observation that distributions of individual blood lead levels are adequately characterized by lognormal distributions. It is especially important to note that the variability over time is distinct from the individual variability in blood lead addressed by the statistical component of the model. In addition, the lognormal variability used by the IEUBK model was estimated as a cross-sectional measure of variability.

Table 1. Results of technical document comparisons relating to the IEUBK model, summarized from the IV&V report.

Technical documents reviewed/compared	Summary of results
Comparison of equations in Appendix A of the TSD to main body of same document.	Discrepancies noted were presented to the U.S. EPA.
Review of the TSD for typographic and editorial errors.	Typographical/editorial errors were rectified by the U.S. EPA.
Review of the Bridge document and comparison with the TSD and with the IEUBK model source code.	Discrepancies were identified between the equations parameters in the Bridge document and the TSD.
	Discrepancies were identified between the location of equations and default parameter values in the IEUBK model source code and the line locations provided in the Bridge document.
	Discrepancies were identified between the equations and default parameter values in the IEUBK model source code and the equations provided in the equation. dictionary of the TSD.
Review of default parameter values in the IEUBK model source code with the Bridge document and the TSD.	The IV&V report identifies parameters in the model source code that have default values but are not discussed in the Bridge document or TSD.
Review of *Draft Final Report for Task 2-8-1: Review Comments on TSD* provided by Battelle.	No contradictions were noted between the reviews.

Table 2. Results from the verification of the default values for the parameters in the IEUBK model source code, summarized from the IV&V report.

Verification activity	Summary of results
Verify parameter values (comparison of default parameter values and parameter units in the IEUBK model source code with those in the TSD).	Several inconsistencies between documentation and the IEUBK model parameter values, parameter names, and a default value were identified.
Review to determine consistency of parameter coding in the source code modules.	To improve the consistency of parameter coding, the IV&V report recommends that all external parameter declarations be made prior to the internal parameter declarations, and that all default value assignments be grouped so that IEUBK model users can easily identify them.
Verify that the assignment of default values to parameters is made in clearly identified blocks of the IEUBK model source code and prior to the use of those parameters.	Very few of the parameters used in the IEUBK model are defined by comment statements in the source code. The IV&V report recommends grouping parameter declarations into general categories and adding a comment block immediately prior to the parameter declarations that defines the group.
Verify that units of measurement are identified in comment statements in the IEUBK model source code.	The units of measurement for default parameter values are not identified in the IEUBK model source code. Units of measure should be identified as part of the comment statement that defines the parameter.

Comments on the Graphics Presented by the IEUBK Model

After reviewing the calculation of the means and the use of internal parameters, each of the IEUBK model subroutines that generate the graphs of the probability distributions were reviewed. The IV&V report concludes that the graphical functions of the IEUBK model work properly.

Table 3. Results for the verification of simple linear equations in the IEUBK model source code, summarized from the IV&V report.

Verification activity	Summary of results
Verify the correct representation in the IEUBK model source code of the equations in the TSD equation dictionary.	The majority of the equations as listed in the TSD equation dictionary are correctly represented in the IEUBK model source code. A few equations were different. However, these differences were either explained in the TSD or were not expected to change the result of the IEUBK model calculation (e.g., typographical errors in the TSD).
Verify the correct call of default values of independent parameters and constants into the equation.	The majority of the equations tested called all of their default parameters correctly.
	The IV&V report recommends that the IEUBK model source code be searched for all occurrences of multiple parameter declarations to ensure that they are consistent. This discrepancy is not expected to significantly affect the results calculated by the IEUBK model.
	Differences in the parameter declaration cause the values used by the IEUBK model to be different at less than the fourth or fifth decimal place and output is reported to the user to the first or second decimal place.
Verify the correct execution of the IEUBK model algorithms and correct reporting of the dependent parameters using Mathcad, version 5.0 (5).	The linear equations execute correctly in the IEUBK model source code.

Table 4. Results of the verification of the execution of differential equations, summarized from the IV&V report.

Verification activity	Summary of results
Verify correct representation in the IEUBK model source code of the difference equations in the TSD equation dictionary.	The majority of equations in the TSD and the IEUBK model source code match.
	Differences do not affect the IEUBK model results.
Verify the correct specification of the 4-hr time interval.	The IEUBK model executes each equation 180 times for each monthly period and therefore is correct.
Verify the correct call of default values of independent parameters and constants into the equation.	All default values are called correctly. Values computed for each difference and associated equation are correct.
Examine use of valid and numerically stable differential equation solution methods in implementing the difference equations.	The backward Euler solution method used by the IEUBK model is a well-tested, widely used method of solving differential equations. No change to this solution method is recommended.

Table 5. Results of the verification of the execution of the overall set of equations, summarized from the IV&V report.

Verification activity	Summary of results
Compare results of the default test scenario for the Mathcad and the IEUBK models.	Several recommendations are provided to improve consistency in the IEUBK model source code to make IEUBK model calculations more accessible to model users.
Compare the results of the multiple test scenarios for the Mathcad and the IEUBK models.	The IV&V report noted that the IEUBK model source code was confusing for three variables used to calculate soil and dust exposures but emphasized that the IEUBK model does execute the exposure equations for soil and dust correctly. The logic of the code is simply hard to follow because it is not consistent.

Model Testing

IEUBK model testing also includes a sensitivity analysis that will address work that has not been addressed in these proceedings. However, several questions have been raised and addressed in response to user questions. For example, one user questioned results showing that uptake from water and from diet decreased as soil lead concentrations were increased. According to the model theory, and as executed in the model source code, media uptake (soil, water, diet, or other) through the gastrointestinal tract is a function of the environmental exposure and of two uptake factors, passive and saturable uptake. The passive, or nonsaturable, uptake is a linear function of exposure and has no limit as environmental concentrations increase. The saturable uptake component is a function of uptake from all media and has an upper limit. Therefore, as uptake from one media (in this case, soil) increases because of increased environmental exposure, uptake from other media (in this case, water and diet) will in fact decrease. Although this is not intuitive at first, it is consistent with model theory. This relationship is shown graphically in Figures 1 and 2.

The test variable used in Figure 1 is derived from the uptake equations in the model equation set U-1 as defined in the TSD. PAF represents the passive absorption factor, $AVINTAKE(t)$ is the total available intake, and $SATINTAKE(t)$ is the half-saturation absorbable lead intake, or the potential uptake of lead when the saturable uptake pathway is 50% saturated. As the diagram shows, the test variable decreases to a constant value as the total available intake increases with increased environmental exposure from any media. Each media uptake is proportional to the lead exposure in media and to the test variable, so that as the test variable decreases, uptake from a particular media will decrease if the exposure to that media is constant. That is, the total available intake will increase as soil lead levels increase, but exposures to diet and water will remain constant. Therefore, the total uptake from diet and water will decrease. This relationship is demonstrated for water in Figure 2.

As shown in Figures 1 and 2, the soil lead levels were increased to soil lead levels well beyond the recommended limits for application of the IEUBK model of <5000 ppm lead in soil. The model results indicated that there would be no upper limit to either the total uptake or the blood lead

level as soil lead levels increased. This would also be true as exposure to any other media was increased. The predicted blood lead levels resulting from each soil lead level are provided in Figure 3. No other parameters were changed during this analysis. Given the U.S. EPA's goal of protecting a typical child or similarly exposed group of children to the 95% confidence limit from a blood lead level of 10 μg/dl, the higher soil and blood lead levels were evaluated only to evaluate IEUBK model performance. An important limitation to the application of the IEUBK model is that it has been designed and should be used to evaluate exposure scenarios that are at least similar to the exposure scenarios used for calibration and empirical validation. In general, the soil lead levels used in these studies are below 5000 ppm lead in soil. Similarly, environmental exposures associated with blood lead levels above 30 μg/dl are above the range of values that have been used in the calibration and empirical validation of this model. Thus, the utility of these figures is to demonstrate the integrity of the mathematical relationships tested.

Discussion

The findings of this work show that the IEUBK model performs calculations in a manner that is consistent with the model documentation. However, this work also shows that the IEUBK model documentation should be improved. The IV&V report recommends that the U.S. EPA revise documentation. The documentation revisions should also address comments in the IEUBK model source code, which do not affect the integrity of computations, but will improve the transparency of computations (9).

The U.S. EPA is undertaking an effort that will address the documentation issues identified in the IV&V work. However, the U.S. EPA also recognizes that there are performance advantages to reprogramming

Figure 1. Values of the test variable as soil lead concentration increases. Note: PAF = 0.2 for all analyses. Theoretically, the limit of the test variable should be equal to the PAF as AVINTAKE(t) increases due to the increase in soil lead concentration. Test variable = PAF + ((1−PAF)/(1+(AVINTAKE(t)/SATINTAKE(t)))).

Figure 2. Lead uptake in water as soil lead concentration increases

Figure 3. Blood lead levels as soil lead concentration increases.

Table 6. Results of the verification of the input/output routines, summarized from the IV&V report.

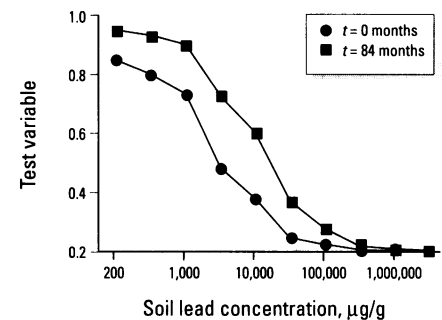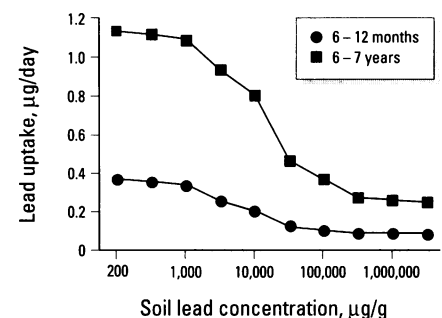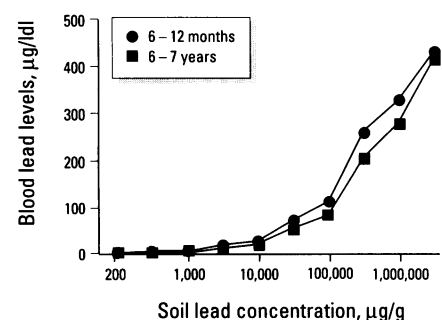| Verification activity	Summary of results
Enter modified parameter values into the IEUBK model and use the C debugger to view how the IEUBK model calls and inputs the modified value.	Regardless of the number of digits (up to six) that a user inputs, the data entry screen on the IEUBK model formats the parameter to between one and three decimal places, but only after the cursor is moved off the data entry field, then back to the same field. The user then may not know which figure (the rounded or the actual figure key entered) is used in IEUBK model calculations. The IV&V report recommends that the IEUBK model source code be revised so that actual values entered are used.
	No quality assurance/quality control checks are coded into the IEUBK model for parameter values that are loaded from existing ASCII data files. The IV&V report recommends that the IEUBK model developers revise the file input routines to include quality assurance/quality control procedures for data loaded from saved ASCII data files.
Analyze the IEUBK model to determine how results that are generated and stored in the calculation routines are communicated to the user or to output data files.	The IEUBK model reuses variable names to improve the readability of the IEUBK model source code. The repeated use of variable names can cause confusion for users attempting to trace function paths in the IEUBK model source code. The IV&V report recommends that identical model variable and parameter names be changed in the IEUBK model source code to clarify the IEUBK model functions. This issue does not affect the results calculated by the IEUBK model.
	Data for gastrointestinal values/bioavailability may be entered from any media data entry screen. It is not made clear that the user is changing the same parameter values regardless of which media the bioavailability data are entered from. This problem does not affect the results of the IEUBK model, but can allow users to construct model scenarios that are different than the user intended. The IV&V report recommends that a single data entry option for these parameters be created in the data entry menu.
Observe the output of results to the screen or to data files, and verify that the results correspond to the values actually generated by the IEUBK model source code.	The parameters PaintConc and PaintFraction are not reported to the user. The statement that is intended to store values for both PaintConc and PaintFraction instead contains the parameter names OccupConc and OccupFraction. This does not affect the results calculated by the IEUBK model, only the input parameters that are reported to the user.

computer code for a Microsoft Windows environment (10). For this reason, the U.S. EPA has initiated work to both address the documentation issues and reprogram the computer code for a Windows environment. The questions from this workshop and other communications with outside scientists and technical experts have also identified a need for more information to better understand the U.S. EPA approach to assessing lead risks for Superfund sites. As such, the technical review work group is developing a lead website. (*http://www.epa.gov/superfund/oerr/ini_pro/lead/index.htm*).

REFERENCES AND NOTES

1. U.S. EPA. Guidance Manual for the Integrated Exposure Uptake Biokinetic Model for Lead in Children. EPA/540/R-93/081. Washington:U.S. Environmental Protection Agency, 1994.
2. U.S. EPA. National Oil and Hazardous Substances Pollution Contingency Plan. CFR 40:Part 300. Washington:U.S. Environmental Protection Agency. 1990
3. U.S. EPA. Revised Interim Soil Lead Guidance for CERCLA Sites and RCRA Corrective Action Facilities. EPA/540/F-94/043. Washington:U.S. Environmental Protection Agency, 1994.
4. U.S. EPA. Validation Strategy for the Integrated Exposure Uptake Biokinetic Model for Lead in Children. EPA/540/R-94/039. Washington:U.S. Environmental Protection Agency, 1994.
5. Mathsoft. Mathcad 5.0 for Windows. Mathsoft, Inc. 101 Main St. Cambridge, MA 02142.
6. U.S. EPA. Report for the Independent Verification and Validation (IV&V) of the Integrated Exposure Uptake Biokinetic (IEUBK) Model for Lead in Children. Report to the Office of Solid Waste and Emergency Response. Research Triangle Park, NC:U.S. Environmental Protection Agency, 1995.
7. U.S. EPA. Technical Support Document: Parameters and Equations Used in the Integrated Exposure Uptake Biokinetic Model for Lead in Children (v. 0.99d). EPA 540/R-94/040. Washington:U.S. Environmental Protection Agency, 1994.
8. Batelle. Correspondence between the IEUBK Lead Model Source Code and Technical Support Document (TSD): Parameters and Equations Used in the Integrated Exposure Uptake Biokinetic Model for Lead in Children (v. 0.99d). Draft Final Report. Research Triangle Park, NC:U.S. Environmental Protection Agency, 1994.
9. U.S. EPA. Report of the Agency Task Force on Environmental Regulatory Modeling: Guidance, Support Needs, Draft Criteria and Charter. EPA 500-R-94-001. Washington:U.S. Environmental Protection Agency, 1994.
10. Microsoft Corp. Windows '95. Microsoft Corp., Redmond, WA 90852.